10-26-2006

The Institutional Repository at the University of Nebraska–Lincoln: Its First Year of Operations

Paul Royster
University of Nebraska-Lincoln, proyster@unl.edu

The Institutional Repository at the University of Nebraska–Lincoln: Its First Year of Operations

The institutional repository (hereafter IR) at the University of Nebraska–Lincoln (hereafter UNL) began operations July 1, 2005. During its first year, it added nearly 2,400 articles and delivered over 94,000 downloads.

The prehistory: The UNL Libraries, led by Dean Joan Giesecke, decided to implement an IR in 2004 and, after a study of its aims and objectives and of the available platforms, selected the Digital Commons software developed by bepress of Berkeley, California, and marketed and licensed through ProQuest, Inc., of Ann Arbor, Michigan. The perceived advantages of this system were 1) its out-of-the-box functionality, 2) its single-price arrangement for software, server space, and ongoing support, and 3) its offering by a known vendor with whom the library had an ongoing relationship.

The first (and still the largest) body of content was the complete set of UNL Ph.D. dissertations since 1955, which UNL contracted with ProQuest to digitize and mount on the site. These were digitized from microfilm copies held at ProQuest/UMI and mounted as page-image PDF files, with most document files running from 10 to 20 Mb. By contract with ProQuest, full-text access to this content is free for UNL users; non-UNL users have free abstracts and 24-page previews, with electronic copies available for purchase for about $30. Approximately 9,400 of these dissertations went online in March 2004; about 800 have been added since, representing new degrees awarded and also extending the backfile to the 1890s, when the first Ph.D.s were granted at UNL. These remained the only content in the IR for the next three months.

I began as the Coordinator for Scholarly Communications in June of 2005, having been brought on staff specifically to manage and recruit content for the new IR. My background of 25 years in scholarly publishing included editorial, production, and copyright experience. At that time, I had been on UNL campus about 2 ½ years, with the English Department and the university press, so I had some faculty contacts and a general sense of the university's organization and its research strengths.

The first several weeks were spent learning the software and creating an architecture for the repository. The Digital Commons software is built around "communities" and "series": documents are held in a series; a series is held by a community; a community can hold multiple series and other communities. I set up "communities" for most of the university's departments and programs. An important decision to be made early on was how to arrange or organize the hierarchy of communities: whether to reflect the institutional structure of the university or the intellectual structure of the content. Because the institutional structure of the university would probably be opaque to users from outside

UNL and is too complex to capture in a single list or table (with some programs belonging to more than one school or college), it was decided 1) to organize the IR by subject, and 2) to list them alphabetically, putting, for example, the Department of Psychology under P, just before the Quilt Studies Center, which is under Q.

Creating this vast (and empty) structure was perhaps a gamble, but one immediate benefit was the necessity of visiting and becoming familiar with the large number of program websites and departmental home-pages—which provided descriptive copy for and were linked to from each program's "community" page on the IR. Days and weeks of surfing your own university's websites is an essential step in integrating the IR with the institution's web presence and digital resources.

With some basic framework in place, and a PowerPoint presentation and handout in tow, the next step was active recruitment. I began to contact department chairs to try to get the opportunity to make presentations to faculty meetings. Simultaneously, I also contacted selected individual faculty of note. At the faculty meetings, I tried to demonstrate the benefits and utility of self-depositing content in the IR; to the select individual faculty, I offered to digitize and put their work online for them. The responses of each group were fairly uniform: The faculty as a group generally thought the IR was a good idea, but almost none were willing to devote the time to self-archive their work. Individuals for whom I offered to do the permissioning, digitizing, and uploading were all willing to participate and were extremely gratified with the results.[1]

As in most fledgling enterprises, the role of luck can hardly be overstated. Within the first two months, I had made contacts that eventually resulted in our most-downloaded item and our most prolific source of content.

The first of these was at the Harold Manter Laboratory of Parasitology at the Nebraska State Museum, whose director, Dr. Scott Gardner, invited me for a tour and discussion in late August. When you enter someone's office and see that they have five computers turned on, it's usually a good sign (if you're looking for digital content). In the Manter Lab—among racks and cases of tens of thousands of pickled worms in glass jars—I noticed a large stack of what appeared to be a manuscript. Upon inquiry, it turned out to be a reference work that had fallen off the publishing track—The Dictionary of Invertebrate Zoology, compiled and written by Dr. Armand Maggenti (University of California–Davis emeritus) and Dr. Gardner. The authors had labored more than 10 years on their project, which included over 13,000 entries and definitions, with etymologies, and it was eagerly anticipated by other researchers in the many fields on which it touched. The work had been accepted for publication by the University of California Press, but in 2004, when the authors were about to send the final manuscript to the publisher, the press cancelled a number of contracts in the life sciences, including this one. Dr. Gardner remarked with a sigh that they had all but given up hope of conventional publication and were considering publishing it online. Seizing the opening, I offered to help and entreated

him to send me the word-processing files. A few days later, I received about 100 WordPerfect files, and began to concatenate, translate, and copy-edit them. About three weeks and 200,000 clicks later, we had an electronic book, which we posted in mid-September 2005. The initial version of the PDF file was about 9 Mb; a subsequent version prepared with Adobe InDesign reduced the size to about 4 Mb. To make faster downloading for users with dial-ups, I also split the dictionary into separate letters, and loaded all the terms into the "Abstract" field to aid online searches. I worried about spending so much time on a single document, but the dean was very supportive, and it proved to be a good investment. The work became an instant online success, averaging over 30 downloads daily from the beginning. It was been downloaded over 12,000 times in our first year, making it by far our most popular work—despite being over 970 pages long.

The second stroke of good fortune was a meeting with the manager of the Internet Center for Wildlife Damage Management, based in UNL's School of Natural Resources. This center was engaged in a project to digitize all the conference proceedings in their field, going back to the early 1960s, but they lacked an easy way to post these files online without heavy html coding and webmaster intervention. The Digital Commons system provided such an easy solution, and the ICWDM began to supply digitized conference proceedings (for the Bird Control Seminars, the Vertebrate Pest Conferences, the Great Plains Wildlife Damage Control Conferences, Bird Strike Committee Proceedings, and others). I split the volumes into individual articles, converted rtf and doc files to PDF, prepared metadata and uploaded files. In the first year, the ICWDM supplied over 800 articles for mounting in the IR, making them by far our most prolific source of content.

A third fortunate event was the suggestion of a departmental administrator in the Department of Chemical and Biomolecular Engineering. She hired a graduate student to gather and upload their faculty's papers, and I trained him in the appropriate procedures. He would go door to door through their department seeking materials to post, and this proved a successful way for that department achieve its aim of increasing their web presence and online reputation.

By the end of one semester, I had come to the realization that voluntary self-archiving was a utopian illusion, and I began to make a new offer to the faculty meetings I addressed. I would still outline the self-archiving features of the system and demonstrate the ease with which articles could be posted, but I would add a "do it for me" option to the "do it yourself" model I had been promoting: "Email me your vita and permission, and I will handle the rest." This produced a substantially healthier response, and one that, by myself, I was hard-pressed to deliver on. Then, after a meeting with faculty from the Department of Physics and Astronomy, I was completely overwhelmed—faculty members sent vita containing over 1500 articles. Fortunately, in our second year, a modest budget for work-study students has allowed us to make considerable inroads in the backlog I had built up over the second semester by making this more generous offer. The ser-

vices we provide include researching copyrights, securing permissions, finding postable digital copies, scanning printed versions, typesetting author's versions, writing metadata, and uploading files. I have found that teen-age undergraduates quickly become capable of performing these tasks that many tenured faculty found too daunting.

Some counter-intuitive lessons:

1. *Campus publishers are not necessarily a rich source of content.* The UNL campus produces literally thousands of publications, including journals, newsletters, and extension publications. Campus publishers, however, have not generally come forward with content for the IR, even though it would provide a free, permanent, and Google-searched way of posting back issues that provide no revenue. I think this is because their staffs are focused on getting out the next issue, and are rather conservative with respect to electronic publication. We hope that this will change as the IR establishes itself as a successful electronic venue.

2. *The department of computer science may want to do it their own way.* My very first department call was on the Department of Computer Science and Engineering. They already hosted their own archive of articles and technical reports, and were not interested in transferring that content to the IR unless a robot could be programmed to transfer it automatically and integrate the IR system with the system they had already engineered.

3. *Librarians are not more interested in IRs than the faculty at large.* Although the UNL Libraries faculty has been exposed to multiple presentations and appeals, their rate of participation (about 10%) is roughly the same as the faculty at large.

4. *Universities work on university time.* Many programs and projects are still considering their options and "thinking about it." Meanwhile, Google indexes the whole web every three weeks, and one billion people are online searching for information.

5. *Dissertations can be very popular content*. In 2006, UNL began requiring Ph.D. candidates to deposit their dissertations electronically, and I agreed to be a resource person for candidates having trouble generating appropriate PDF files. I wound up helping about 30 candidates, and to each one I explained that their deposit agreement with ProQuest was non-exclusive, giving them the right to post their dissertation in the open-access portion of the IR in addition to the for-pay section run by ProQuest. All but two took advantage of this, and these have become extremely popular items. Such posting also gives hiring committees free and instant desktop access, certainly an advantage to the job candidate. Where a popular for-pay dissertation may be downloaded dozens of times, the free-access ones are downloaded hundreds of times. Last month (August 2006), two of our top five downloads were recent dissertations.

6. *Some faculty are wary of online publication and reluctant to publish*. Working in print publishing, I occasionally encountered an author who came up with many last-minute revisions and reasons to postpone publication—"inkophobia," I called it. "If my precious work is out there, someone may find fault with it"—is a common and understandable anxiety; it very well may be either attacked or (worse) ignored. Add to this the relatively lower status of online publication and the fear of being plagiarized (is there a deeper form of flattery?). I think the greatest bar is the unvoiced fear that no one will notice or care.

Feedback and reinforcement

One of the most useful features of the Digital Commons software has been the monthly reports of downloads that are automatically sent to authors. When faculty realize that people are actually downloading their papers, they become much more enthusiastic and often come forward with more obscure and difficult-to-find content. This has been a very important feature, both for recruiting content and for increasing faculty satisfaction.

Publicizing the content and increasing usage

Just as putting up the repository doesn't guarantee that content will come, so also putting up the content doesn't guarantee that usage and downloads will follow. In many cases it does, thanks to Google searches or referring websites, but there are ways that usage can be encouraged and promoted. We have also begun cataloging book-length works and complete run serials in our regular library catalogue, which has the advantage of exposing them to the WorldCat system, as Internet Resources. (A list of our twn most-downloaded documents and a chart showing downloads by month from the IR are included in the appendices.)

The success of the *Online Dictionary of Invertebrate Zoology* encouraged me to do more online promotion, especially when I had developed some of my own publications whose usage I wished to increase. The best sites are those that are active and continually make updates. **Wikipedia** (http://en.wikipedia.org) is an excellent place to put referring links. The site gets a good deal of traffic and links can be written in and become active immediately. The **Online Books Page** (http://onlinebooks.library.upenn.edu/) at University of Pennsylvania is an excellent referring site for book-length works. It may take several weeks for links to submitted titles to appear, but usage increases dramatically when they do. Other sites, such as HUMBUL, MERLOT, TAMU, ERIC, and the Internet Public Library, have proved less hospitable—having specific requirements for content or file formats or hosting mechanisms that don't jive with our IR's architecture and content mix.

Some sites claim to be the latest word in providing links and information, until one realizes that they have not updated since the Internet bust of 1999. Sometimes the "informa-

tion highway" reminds me of R. W. Emerson's description of a Western road—that it starts out as a grand tree-lined royal boulevard, then becomes a regular street, then a winding rutted country lane, then a foot-path, and then turns into a squirrel-track and runs up a tree.

Final Summary

At the close of our first year we had mounted 2,397 articles, which had been downloaded 58,743 times; the dissertations collection included 10,117 theses, which had been downloaded 34,862 times. We currently average around 2,000 hits and 700 downloads per day. Usage increases as the semester progresses, and falls during the Christmas break and during the summer.

In addition to the various wildlife damage conference proceedings, we had mounted extensive backfiles of the *Journal of Parasitology*, and the proceedings of the People of Color at Predominantly White Institutions conferences. Over 400 faculty members were represented as authors; and we had assembled significant collections of materials in physics, astronomy, textiles, quilt studies, psychology, family and consumer science, agricultural economics, English, modern languages, classics, religious studies, natural resources, mammalogy, entomology, ornithology, invertebrate zoology, materials and nanoscience, architecture, community and regional planning, and chemical engineering. We also have materials in women's studies, mechanical engineering, music, political science, journalism, mathematics, chemistry, history, geoscience, and zoonotics.

On a personal note, I have found it extremely rewarding to engage with scholars across a very wide range of subjects and disciplines. I have gained a passing acquaintance with such diverse subjects as zoological nomenclature, Roman concrete, bat dentition, Cronbach's α, chronic wasting disease, lesbian pedagogy, fabric dyes and preservation, prosocial behavior, biodiesel processing, Eskimo curlews, scatological French poetry, nanomagnetics, and variable stars. I have also found an outlet for publishing my own personal scholarly endeavors (digital editions of early American primary works). I have no hopes of ever catching up with the campus's scholarly output, but I look forward to staying busy and doing something that people appreciate. It's hard to ask for more than that.

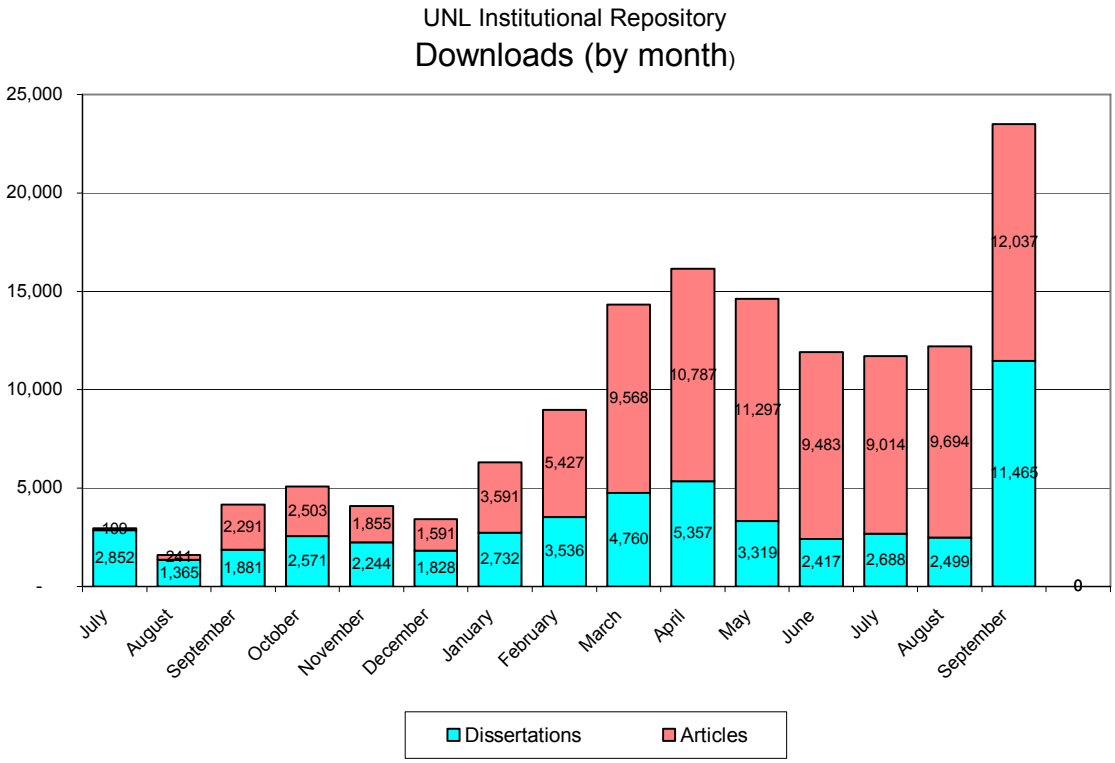
Paul Royster
Coordinator of Scholarly Communications
University Libraries, University of Nebraska–Lincoln

Appendix 1

The most popular downloads: July 1, 2005–June 30, 2006

	downloads
1. Online Dictionary of Invertebrate Zoology http://digitalcommons.unl.edu/onlinedictinvertzoology	12,100
2. The Constitutions of the Free-Masons (1734). An Online Electronic Edition http://digitalcommons.unl.edu/libraryscience/25	2,050
3. Thomas Pynchon: A Brief Chronology http://digitalcommons.unl.edu/libraryscience/2	759
4. Improved Conversion of Plant Oils and Animal Fats into Biodiesel and Co-Product http://digitalcommons.unl.edu/chemeng_biomaterials/20	755
5. Immigration, the American West, and the Twentieth Century: German from Russia, Omaha Indian, and Vietnamese-Urban Villagers in Lincoln, Nebraska [a Ph.D. dissertation] http://digitalcommons.unl.edu/historydiss/1	657
6. Process for Producing Bbiodiesel Fuel with Reduced Viscosity and a Cloud Point below Thirty-two (32) Degrees Fahrenheit http://digitalcommons.unl.edu/chemeng_biomaterials/1	615
7. The Christian Commonwealth: or,The Civil Policy Of The Rising Kingdom of Jesus Christ (1659). An Online Electronic Text Edition http://digitalcommons.unl.edu/libraryscience/19	504
8. David Cusick's Sketches of Ancient History of the Six Nations (1828) http://digitalcommons.unl.edu/libraryscience/24	475
9. Production of Ethers of Glycerol from Crude Glycerol—The By-Product of Biodiesel Production http://digitalcommons.unl.edu/chemeng_biomaterials/18	461
10. A Brief History of the Warr with the Indians in New-England (1676): An Online Electronic Text Edition http://digitalcommons.unl.edu/libraryscience/31	446

Appendix 2



[1] I am indebted to Dr. Patricia Crews, Dr. Paul Johnsgard, and Dr. Carolyn Pope Edwards for their indulgence, support, and cooperation in this effort.